

comparing the two automated methods of model building from poor initial data.

5 The second set of RMSD values is for the present lattice models, which for a convenient comparison were converted into the full-atom models also via an automatic application of MODELLER (with lattice models of the C α backbones used as templates). As indicated, the most significant improvement of the model quality occurs when the threading alignment produces a rather poor but not
10 nonsensical initial model (compare Tables X and XI). As shown in Table XI, for small globular proteins, such threading-based models have an RMSD in the range of 6-8 Å from native (over the aligned fragments). When the threading models are poor, *e.g.*, for lcewI or 2azaA, the improvement is rather small. At the other extreme are those cases when the alignment is good, and the resulting RMSD relatively
15 small. Here also, the changes are small because the models are already good. Importantly, the procedure essentially does no harm to these models; thus, it can be applied to all situations with impunity. In summary, in 6 of 9 test cases (in 9 of 12 including the three proteins employed in the model turning procedure), the models generated by the invention give lower values of RMSD over the set of aligned
20 residues. In the three remaining cases, the changes in RMSD were insignificant (essentially in the range of the statistical fluctuations). In five cases, qualitative improvements were observed (for the aligned residues as well as for entire models; compare data given in Table 4): from 5.6 Å to 3.5 Å for lhom, from 7.1 Å to 4.7 Å for 1stfl, from 7.9 Å to 3.9 Å for 1tlk, from 6.9 Å to 4.4 Å for 256b or from 6.6 Å to
25 4.4 Å for 2pcy. These numbers were for the initial threading and final lattice (refined with MODELLER) models, respectively. It should be noted that the MODELLER refinement of the final lattice models changed their RMSD very little (in the range of 0.2 Å), while the improvement of the initial threading models by the application of MODELLER was more noticeable.

30 It is very interesting to see how the proposed procedure deals with the non-aligned part of the model. Comparison of the RMSD values for the aligned parts (Table XI) and for the entire structured parts (Table X) of the model reveals that the

algorithm built rather reasonable models of that entire structure, provided there was a well defined fragment of good geometrical fidelity in the original alignment. Again, in all but two cases, the present method lead to more accurate models. For both the aligned part of the molecules and for entire chains (Table 4), good models were generated in about half of the studied cases (including all three proteins used in the model turning procedure). In the remaining cases, models were seen that were marginally improved, as for 3cd4, or that remained rather poor final models, as for 2azaA or 5fdl; this was true despite an RMSD decrease of more than 10 Å, as compared to models generated automatically by MODELLER from the initial threading results.

Discussion

Means of the model improvement

There are several ways in which the invention changes the protein model from the original fragmentary threading model. First, non-aligned parts (*e.g.*, loops) are added and readjusted according, to packing requirements and the preferences encoded in the force field. Then, the entire chain has some freedom of movement within the template tube without any changes in its template-target sequence assignment. Furthermore, parts of the chain can slide along the tube, thereby allowing for a quite substantial modification of the initial alignment and, consequently, the resulting structure. Finally, the aligned fragments can leave the tube in a lateral direction. These segments can enter a different part of the template tube or remain outside of it. Such motions of the model chain could result in a large change of the structures, or even a change of the fold topology. The last, rather radical mode of the model rearrangements happened in several cases. In other words, the most effective way of model improvement was by neglecting a part of the threading alignment, even at the expense of various template-related energetical penalties. Interestingly, those sections of the threading-based model that were

consistent with the target structure underwent only very minor changes in all cases, and the alignment remained unchanged. As discussed below, this observation may help identify those models that should be of good quality from those for which improvement of the starting threading model is not satisfactory.

Below, for three selected cases, more detailed specific rearrangements of the initial threading models that took place during the Monte Carlo simulations are presented.

2pcy

The threading alignment of the 2pcy sequence on 2azaA covered a substantial part of the sequence. There are gaps of substantial length. As a result, the threading model had the wrong topology, and two-edge strands of the eight-member β -barrel (one in each of the two P-sheets) were located in the wrong sheets. This was the reason for the resulting 7.6 Å RMSD from native for the models built solely from the threading alignment. During the simulations, the three C-terminal strands remained almost unchanged. Similarly, the three N-terminal strands underwent only small adjustments; however, in several models, one or two strands slid along the tube by a distance that sometimes changed the original alignment by one or two positions. The central fragment of the model chain (two putative irregular strands, with a couple of short helices breaking these strands) was responsible for the large RMSD in the initial model. The algorithm erased most of the template-target assignments in this part of molecule. Partly this occurred because of the compactness criterion; several residues did not have any long-range contacts in the threading model. During the simulated annealing process, residues 30-37 (small differences in the extension of this fragment can be seen between the particular runs) switched their sheet assignment, and joined the tube fragment associated with one of the C-terminal β -strands, the third one from the C-terminus. This was seen in the final "new assignment", or pseudo-alignment. At the same time, the second strand (completely helical in the threading model) moved to the